BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 5: e74114
doi: 10.3897/biss.5.74114

OPEN ACCESS

Conference Abstract

Algorithms for connecting scientific names with literature in the Biodiversity Heritage Library via the Global Names Project and Catalogue of Life

Geoffrey D Ower‡, Dmitry Mozzherin‡

‡ University of Illinois, Champaign, United States of America

Corresponding author: Geoffrey D Ower (gdower@illinois.edu), Dmitry Mozzherin (dmozzherin@gmail.com)

Received: 07 Sep 2021 | Published: 08 Sep 2021

Citation: Ower GD, Mozzherin D (2021) Algorithms for connecting scientific names with literature in the Biodiversity Heritage Library via the Global Names Project and Catalogue of Life. Biodiversity Information Science and Standards 5: e74114. https://doi.org/10.3897/biss.5.74114

Abstract

Being able to quickly find and access original species descriptions is essential for efficiently conducting taxonomic research. Linking scientific name queries to the original species description is challenging and requires taxonomic intelligence because on average there are an estimated three scientific names associated with each currently accepted species, and many historical scientific names have fallen into disuse from being synonymized or forgotten. Additionally, non-standard usage of journal abbreviations can make it difficult to automatically disambiguate bibliographic citations and ascribe them to the correct publication. The largest open access resource for biodiversity literature is the Biodiversity Heritage Library (BHL), which was built by a consortium of natural history institutions and contains over 200,000 digitized volumes of natural history publications spanning hundreds of years of biological research. Catalogue of Life (CoL) is the largest aggregator of scientific names globally, publishing an annual checklist of currently accepted scientific names and their historical synonyms. TaxonWorks is an integrative web-based workbench that facilitates collaboration on biodiversity informatics research between scientists and developers. The Global Names project has been collaborating with BHL, TaxonWorks, and CoL to develop a Global Names Index that links all of these services together by finding scientific names in BHL and using the taxonomic intelligence provided by CoL to conveniently link directly to the page referenced in BHL. The Global Names Index is

© Ower G, Mozzherin D. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

continuously updated as metadata is improved and digitization technologies advance to provide more accurate optical character recognition (OCR) of scanned texts. We developed an open source tool, "BHLnames," and launched a restful application programming interface (API) service with a freely available Javascript widget that can be embedded on any website to link scientific names to literature citations in BHL. If no bibliographic citation is provided, the widget will link to the oldest name usage in BHL, which often is the original species description. The BHLnames widget can also be used to browse all mentions of a scientific name and its synonyms in BHL, which could make the tool more broadly useful for studying the natural history of any species.

Keywords

biodiversity informatics, scientific names index

Presenting author

Geoffrey D. Ower

Presented at

TDWG 2021